(4/25/2017)

Phylogenetic Profilers

Genes that may require further analysis and functional annotation can be identified using comparative analysis tools within a specific genomic context. Phylogenetic profiles are designed for such purpose.

Single Genes

In many cases the differences in physiology, phenotypic properties and ecology of different organisms can be attributed to the differences in their gene content, i.e., the differences in abundance of various gene families, including the ultimate case of certain genes being present in one genome but not in another genome(s) and vice versa. Therefore the genes identified as more or less abundant (or present or absent) when comparing the genome of interest to its genome context, often become the focus of microbial genome analysis and may require special attention from the annotator.

The Phylogenetic Profiler for Single Genes tool allows finding genes in a specific genome that have / do not have homologs in other related genomes. There are two steps involved in such a selection:

(a) Start with Find Genes in the Main Menu and select Phylogenetic Profiler for Single Genes under the Phylogenetic Profilers second level menu bar.

(b) Select a target genome and set the condition for selecting its genes with respect to presence or absence of homologs in other related genomes.

(c) (Optional) Additional features are available for setting cutoffs and adding extra fields in the output display.

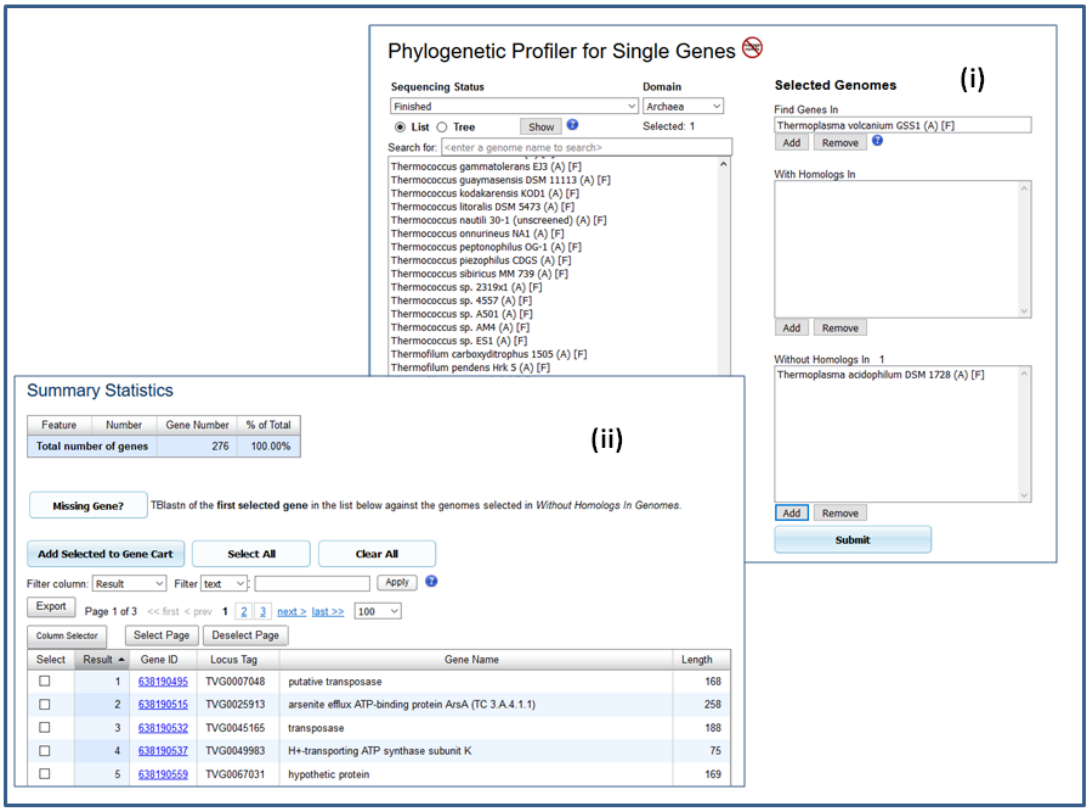

Figure 1. Find genes with the Phylogenetic Profiler for Single Genes tool.

Example 1. After setting the genome context to two genomes, T. volcanium (*Thermoplasma volcanium GSS1*) and T. acidophilum (*Thermoplasma acidophilum DSM 1728*), use the Phylogenetic Profiler to find T. volcanium genes that have no homologs in T. acidophilum, as shown in Figure 1(i). Similarity cutoffs can be used to fine-tune the selection. To choose your own cutoff values, scroll down to find **Similarity Cutoffs** under **Advance Options**. The list of genes with the specified profile are then provided as a selectable list as shown in Figure 1(ii). Users can also select to add additional fields such as COG, Pfam, etc. in the result display. Such selections are available in the **Function Display Options** under the same **Advance Options** in the window shown in Figure 1(i).

The Phylogenetic Profiler for Single Genes can be used, for example for finding unique, conserved, or gained genes in the target genome with respect to other genomes of interest. In the example shown in Figure 1, 276 genes are found to be unique in T. volcanium with respect to T. acidophilum. These genes can be selected to add into the Gene Cart for further analysis.

Gene Cassettes

The Phylogenetic Profiler for Gene Cassettes allows selecting genes that are part of a gene cassette (i.e., are collocated on the choromosome) in a query genome and are part of related (conserved part of) gene

cassettes in other genomes. (**Limitation:** Currently you can only select up to **50 Collocated In** Genomes.)

To use this tool, first, select the protein cluster you wish to use in the comparison (COG or Pfam). Then select your query genome into the "Find Genes In" column, and select (up to 50) genomes for gene cassette comparisons with the query genome in the "Collocated In" list. Click "Submit" button to start the analysis.

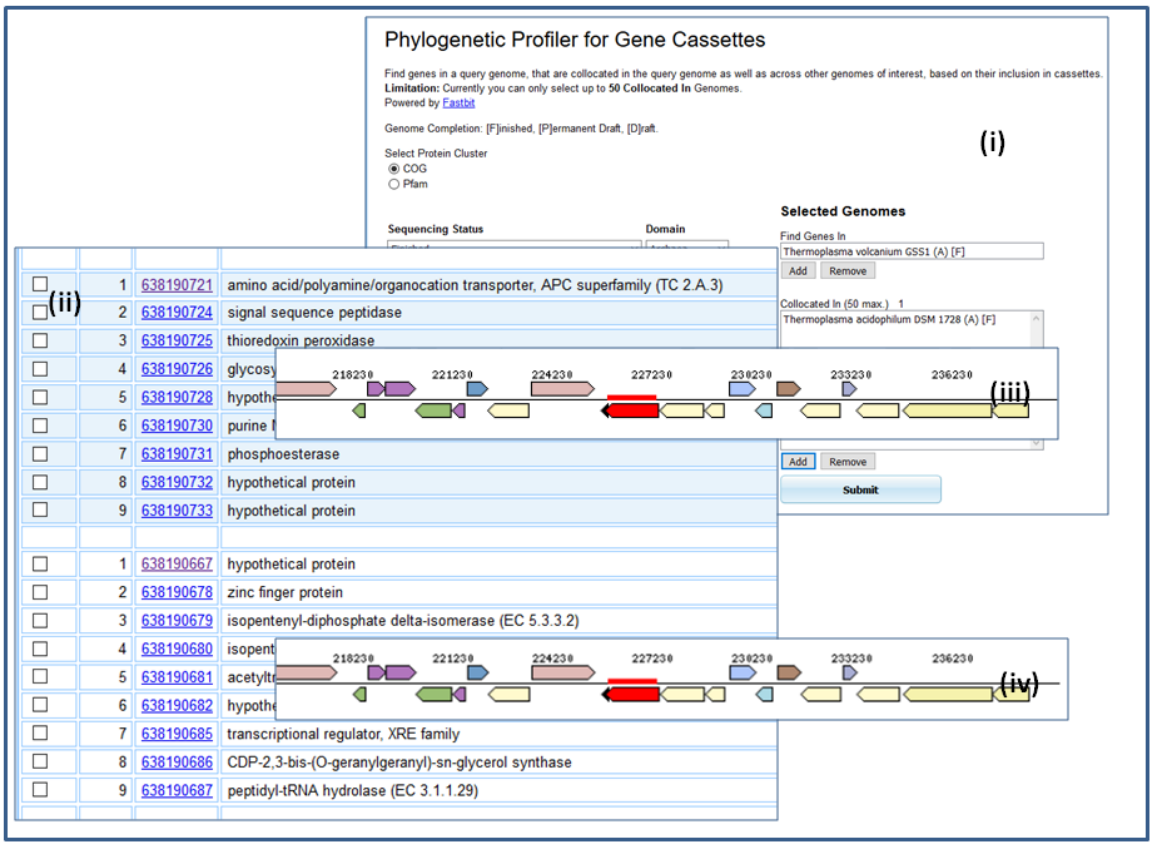

Figure 2. Find genes with the Phylogenetic Profiler for Gene Cassettes tool.

Example 2. Suppose a user wishes to find gene cassettes in T. volcanium (*Thermoplasma volcanium GSS1*) collocated in T. acidophilum (*Thermoplasma acidophilum DSM 1728*) using COG Protein Cluster. T. volcanium will be added to the "Find Genes In" field, and T. acidophilum will be added to the "Collocated In" list as shown in Figure 2(i).

After the user clicks the Submit button, the result will show all collocated gene cassettes in both genomes as in Figure 2(ii). There are two corresponding 9-gene cassettes in T. volcanium (starting from gene 638190721) and in T. acidophilum (starting from gene 638190667), respectively. Chromosome viewer -- colored by COG for both groups are shown in Figure 2(iii) and 2(iv). (To see the chromosome viewer colored by COG, first go to the gene detail page by clicking the Gene OID, then select "COG" in

the "Chromosome Viewer colored by" dropdown list in the **Evidence For Function Prediction** section in the gene detail page.)

Note that in each specific group of collocated genes in the query genome, individual genes may correspond to parts of multiple chromosomal cassettes in the other genomes involved in the profiler condition. Users can explore the details of individual gene listed in the result by clicking on the associated Gene OIDs. Users can also select all or a subset of genes from the list to be added to the Gene Cart for further analysis.